

DOCUMENT RESUME

ED 201 668

TM 810 262

AUTHOR Hendrickson, Les
TITLE A Secondary Analysis of Philadelphia Reading Achievement Using Longitudinal Causal Models. Part One, Draft.
PUB DATE 23 Apr 81
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981). Tables in small print.
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Gains; *Cohort Analysis; Comparative Analysis; Elementary Education; Error of Measurement; Factor Structure; Grade 3; Grade 4; *Longitudinal Studies; *Mathematical Models; Multiple Regression Analysis; *Reading Achievement
IDENTIFIERS *Causal Models; LISREL Computer Program; *Philadelphia School District PA; Secondary Analysis; Structural Equation Models

ABSTRACT

The study of the same people at several occasions is known by various names: panel studies, multi-wave, multi-variable models, cohort analysis, and gain score analysis. The use of econometric concepts, path analysis, and structural equation analysis to study multi-wave, multi-variable models became widespread in methodological practice in the late 1960's. This paper examines alternative procedures for analyzing a two-wave, one-variable model. The situation studied occurs when a dependent variable and a set of independent variables are measured on the same group of people at two times. The use of multiple regression and structural equation modeling are compared in models with and without measurement error. Two different strategies for studying dependent variables are considered: first, the use of the difference or gain score as a single dependent variable, and secondly, the recursive use of the Time 1 and Time 2 data together. The LISREL structural equation approach is used to analyze data which resulted from a joint effort in 1977 and 1978 of the Federal Reserve Bank and the School District of Philadelphia to study factors affecting the reading achievement of 1,800 elementary school children. This data was reported in "What Works in Reading." (Author/RL)

DRAFT

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

A SECONDARY ANALYSIS OF PHILADELPHIA READING ACHIEVEMENT
USING LONGITUDINAL CAUSAL MODELS

PART ONE

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

L. Hendrickson

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Les Hendrickson
Research, Analysis and
Planning Division
Eugene Public Schools
200 North Monroe Street
Eugene, Oregon 97402

A paper presented at the
1981 American Educational
Research Association Meetings
Los Angeles, California
April 23, 1981

INTRODUCTION

The study of the same people at several occasions is known by various names: panel studies, multi-wave, multi-variable models, cohort analysis, and gain score analysis. The use of econometric concepts, path analysis, and structural equation analysis to study multi-wave, multi-variable models became widespread in methodological practice since the late 1960's (Duncan, 1966, 1969; Costner, 1969; Heise, 1969; Land, 1969), after developmental work had been done earlier (Simon, 1957; Blalock, 1964; Boudon, 1965).

The purpose of this paper is to examine alternative procedures for analyzing a two-wave, one-variable model. The situation studied occurs when a dependent variable and a set of independent variables are measured on the same group of people at two times. The use of multiple regression and structural equation modeling are compared in models with and without measurement error. Two different strategies for studying dependent variables are considered: first, the use of the difference or gain score as a single dependent variable, and secondly, the recursive use of the Time 1 and Time 2 data together. The LISREL structural equation approach is used to analyze the data. Pendleton, Warren, and Chang (1975), point out the widespread agreement in sociology and demography that the use of longitudinal models is necessary and increasing.

The Data Base

The data base examined in this study is the result of a joint effort in 1977 and 1978 of the Federal Reserve Bank and the School District of Philadelphia to study factors affecting the reading achievement of 1,800

elementary school children. Approximately 8,000 copies of their report "WHAT WORKS IN READING" and its summary have been distributed throughout the world (Kean, Summers, Raivetz and Farber, 1979).

Information on 25 reading teachers, 25 principals, 94 teachers, 68 reading aides, and the 1,800 students yielded 245 variables which were analyzed using multiple regression techniques. In the Philadelphia study, the sample selection process was done by school. Elementary schools were ranked according to their 1974 and 1975 average Total Reading Achievement Development Scale Scores (ADSS) on the California Achievement Tests (CAT-70) summed over grades 1-4 for all 190 district elementary schools for each of the two years. The schools were ranked on the difference of these sums. The final sample contained ten "high-high", five "middle-middle", and ten "low-low" schools which gave representation from all eight administrative sub-districts of the city. The sample totaled 25 schools. The students in these schools were representative of students in schools having high, middle, and low success in reading achievement. Data collection procedures included interviews with school personnel (e.g., principals, teachers, reading aides) and recording data from pupil records. The data collection process was completed in two weeks, (Kean, Summers, Raivetz and Parker, 1979).

Over 500 multiple regression runs were done to establish which of the 245 variables measured had the most impact on the reading achievement gain. Eighteen variables were identified as contributing to achievement gain.¹ "WHAT WORKS IN READING?" points out the difficulties of analyzing a complex data base without a theory. There is no research tradition in the public school system which cumulatively reports and analyzes the results of large district-wide evaluations of reading instruction. Few school districts have published work comparable to the School District of Philadelphia.

The Philadelphia technical report consists almost entirely of descriptions of variables and correlation matrices. No cross-tabulations, path analysis, or modeling of the 245 variables was undertaken. For example, neither the relations among the 18 variables nor their impact on third and fourth grade test scores are analyzed.

Below, the methods and findings of the gain score analysis used in "WHAT WORKS IN READING?" are briefly discussed. The use of model testing and LISREL is discussed next, then the variables to be analyzed are described, followed by comparisons among the multiple regression results and the results obtained using both correlation and covariance matrices. Measurement error assumptions in two-wave, one-variable models are discussed in the context of identification problems. Part Two of the paper examines alternative models of the Philadelphia data using a sensitivity analysis (Land and Felson, 1978), to look at the reasonableness of common assumptions made in analyzing this class of models.

CHOICE OF DATA ANALYSIS APPROACHES

The Gain Score Analysis

The regression runs done by Federal Reserve Bank economists used the difference between the third and fourth grade reading achievement scores as a single dependent variable. The use of "difference", "change", or "gain" scores has been thoroughly examined (Thorndike and Hagen, 1955; Thorndike, 1966; Bohrnstedt, 1969; Cronbach and Furby, 1970; Alwin and Sullivan, 1975; Kessler, 1977; Pendleton, Warren and Chang, 1979). This examination has generally advised against using gain scores because: the difference between

the two measures has lower reliability than the measures considered separately; their use requires low error variance and high reliability in variables; the calculation of the gain reliability score tends to be unstable because it depends on five other values--the three correlations and two variances; and the analysis of gain scores is complicated by regression toward the mean effects.

Thorndike (1963:40, 1966:124), points out two additional characteristics of using gain scores. First, the gain score is almost certain to be negatively correlated with the initial achievement score. Second, the variance of the gain scores is in some cases no more than one-fourth the size of the variance of the Time 1 and Time 2. Despite these disadvantages, gain scores continue to be used in applied research and evaluation work (Alwin and Sullivan, 1975).

Model Testing and the Use of LISREL

It is theoretically reasonable that the gain score analysis can be extended by considering alternative models that use data from both times rather than using the difference score as a single dependent variable. The analysis below uses the maximum likelihood exploratory factor analysis (EFAP) and structural equation approaches (LISREL) of Jöreskog and Sörbom (1978).

These approaches were introduced to sociologists in the early 1970s (Hauser and Goldberger, 1971; Burt, 1973). Maximum likelihood estimation procedures, originally developed by the British statistician R. A. Fisher (1921), yield estimations which are efficient and consistent for large samples.²

Two-, three-, and four-wave multi-variable models have been extensively studied. See, for example, Duncan, 1972; Hannan and Young, 1977; Hargens, Reskin and Allison, 1976; Jöreskog and Sörbom, 1977; and Wheaton, Muthén, Alvin and Summers, 1977.

Since the early 1970s, substantial experience with LISREL structural equation and related approaches has been published. These approaches allow an extraordinary flexibility of model testing. The late 1970s has seen a period of reflective consideration during which analysis difficulties with these approaches accumulated, assumptions were examined, and procedures developed to deal with the difficulties. Illustrative of this is Fink and Mabee, 1978; Saris, Pijper and Zegwaart, 1979; Gillespie and Fox, 1980; Horn and McArdle, 1980; Schmidt and Schiefley, 1980. A specific example of this process is the finding of negative variances greater than one in the error matrices of identified models. A study of these situations, as Fink and Mabee (1978) show, has lead to the general awareness of the possibility of multiple solutions of model parameters for identified models. Land and Felson (1978), use the term "sensitivity analysis" to describe systematic procedures for testing the effect of alternative parameter assumptions. Gillespie and Fox (1980) exemplify this testing in their study of bias effects in non-recursive model parameter estimations.

DATA ANALYSIS

Descriptions of the Variables and the Correlation Matrix

As mentioned above, "WHAT WORKS IN READING?" found 18 of the 245 variables studied had a statistically significant beta weight in predicting the gain score. Table 1 lists the Philadelphia code names, definitions,

means, and standard deviations for 16 of the independent variables,³ and for the three dependent variables: the gain score, and the third and fourth grade reading scores. The 16 independent variables include measures of student, teacher, principal, and school organization. Seven of the 16 are dichotomous.

TABLE 1
Code Names, Definitions, Means, and Standard Deviations for Sixteen Significant Independent
Variables and Three Dependent Variables

Code Names	Definitions	Mean	Standard Deviation
CATGAIN	Difference between Grade 3 and Grade 4 scale score	28.43	52.50
KCATCGR3	California Achievement Test--Reading Comprehension Scale Score for Grade 3, 1975	385.06	67.74
KDAYPRES	Days Students were present	130.51	10.41
KUNEXABS	Unexcused absences	2.64	6.44
KINDERGN	Student attended kindergarten 1=NO, 2=YES	1.80	0.40
SUPSTFPK	Number of non-teaching supportive staff per school	----*	11.02
XPCTGT84	Percent of students scoring above 84th percentile California Achievement Test 1976--Total Reading	.20	.13
XPCRTL2	Percent of classroom teachers with less than 2 years experience	.197	.136
SGRDORG7	School has kindergarten to seventh grade 1=NO, 2=YES	----*	.46
TPPNOABS	Number of teacher pay periods with no absence	13.89	3.79
TATTCONF	Teacher attends outside professional conference meetings 1=NO, 2=YES	1.17	.39
T1STGR04	First year teaching grade 4 1=NO, 2=YES	1.17	.35
PBRDGEXP	Principal has prior reading experience 1=NO, 2=YES	1.36	.47
TPROBSRP	Times principal observed during reading class	4.41	11.98
TMINRDGM	Minutes per week of individual independent reading	73.35	60.31
TSELROGM	Teacher would select the same reading program again	1.54	.50
T55LNBSL	Teacher uses Basal Linguistic reading program 1=NO, 2=YES	---*	.29
THAIDEX	Times per week aide in room during reading	2.55	2.31
KCATCGR4	California Achievement Test, Reading Comprehension Scale Score for Grade 4, 1976	412.50	72.56

*Means for SUPSTFPK, SGRDORG7, and T55LNBSL were not shown in the November 1979 technical report of "WHAT WORKS IN READING?"

Table 2 shows the 19x19 correlation matrix of the variables described in Table 1. The impression obtained is that the matrix is thin. For example, of the 171 correlations in it, 30 or 18%, are greater than .15, and 18 or 11%, are greater than .25.

Table 2 contains 120 correlations among pairs of the 16 independent variables and 11, or 9%, are greater than .25. The highest correlation of any variable with CATGAIN, the gain score, is .11.

Table 3 shows the variance-covariance matrix of CATGAIN and the third and fourth grade achievement scores. Thorndike's comments (see Thorndike, 1963:40, 1966:124) mentioned earlier, are appropriate here. As anticipated, the gain score is negatively correlated with the initial score and the covariance of the gain score is 60% the size of Time 1, the initial score's covariance and 52% the size of the Time 2 covariance.

TABLE 2

Correlation Matrix for 19 Variables in Philadelphia Reading Achievement Study, N = 1,363

	CATGAIN	KCATCGR3	KDAYPRES	KUNEXABS	KINDERGN	SUPSTFPK	XPCTGT84	XPCRTL2	SGRDORG7	TPPNOABS	TATTCOMF	T1STGR04	PBRUGEXP	TBROBSRP	TMINDRG	TSELROGM	TSSLNBSL	TNAIDWK	KCATCGR4
CATGAIN	1.000																		
KCATCGR3	-.292	1.000																	
KDAYPRES	.074	.161	1.000																
KUNEXABS	-.048	-.239	-.632	1.000															
KINDERGN	.020	.123	.116	-.142	1.000														
SUPSTFPK	-.004	-.290	-.134	.213	-.145	1.000													
XPCTGT84	-.051	.386	.134	-.211	.141	-.626	1.000												
XPCRTL2	.071	-.118	-.056	.119	-.095	.327	-.154	1.000											
SGRDORG7	.029	-.073	-.025	.059	-.042	.254	-.236	.346	1.000										
TPPNOABS	.042	.076	-.017	-.016	-.006	.135	-.040	.078	.247	1.000									
TATTCOMF	-.037	-.112	-.033	-.030	.004	-.027	-.082	.118	-.130	.005	1.000								
T1STGR04	-.002	-.139	-.061	.013	.013	.104	.007	.021	-.045	-.085	.021	1.000							
PBRUGEXP	.112	-.032	.101	-.076	.057	.009	.068	.272	.111	-.086	.000	.219	1.000						
TBROBSRP	-.040	.018	.002	.014	-.007	-.118	-.104	-.218	-.021	-.024	-.075	.051	-.139	1.000					
TMINDRG	.083	.090	.086	-.075	.034	-.143	.113	.005	.069	-.004	-.125	.021	.279	-.075	1.000				
TSELROGM	-.007	.187	-.029	-.006	-.006	-.142	.030	-.033	.073	.080	.023	-.011	-.093	.126	-.119	1.000			
TSSLNBSL	.105	-.088	-.036	.090	-.077	.093	-.180	.224	.004	.013	-.062	.157	.041	-.068	-.088	.112	1.000		
TNAIDWK	.004	-.362	-.116	.139	-.110	.527	-.427	.074	.072	.010	.254	-.012	-.141	.111	-.069	-.136	.011	1.000	
KCATCGR4	N.A.*	.722	.197	-.257	.129	-.273	.382	-.065	-.047	.101	-.132	-.132	-.051	-.012	.144	.169	-.006	1.335	1.000

*The correlation between KCATCGR4 and CATGAIN was not calculated.

TABLE 2

TABLE 3
 Variance-Covariance Matrix of Gain Score with the Two Achievement Scores
 N = 1,363

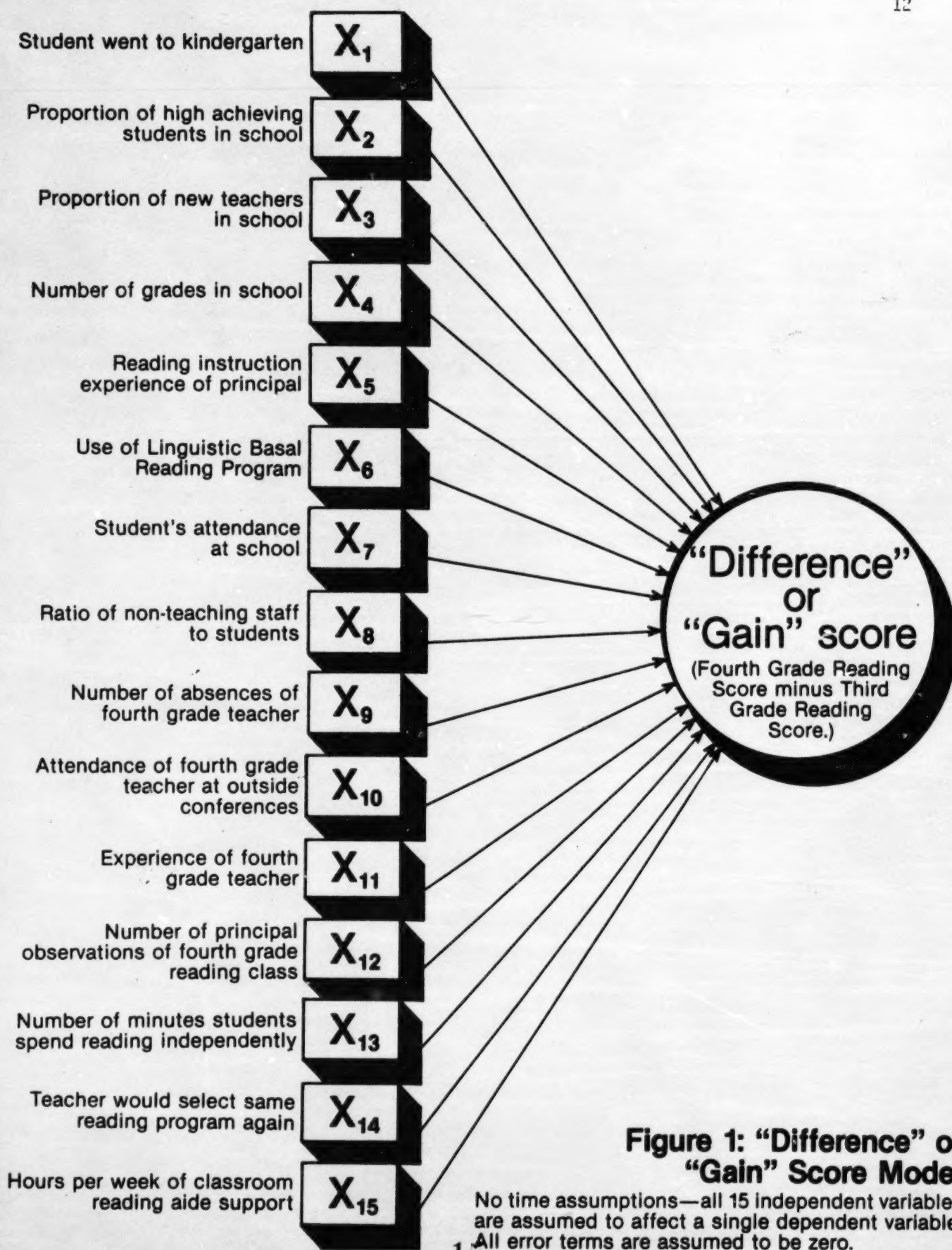
	<u>CATGAIN*</u>	<u>KCATGRD3</u>	<u>KCATGRD4</u>
CATGAIN	2,756.25	-----	-----
KCATGRD3	-1,039.52	4,588.71	-----
KCATGRD4	Not calculated	3,548.78	5,264.95

*Where CATGAIN is the gain score or difference between KCATGRD3 and KCATGRD4 and KCATGRD3 is the student's third grade reading achievement score and KCATGRD4 is the student's fourth grade reading score.

A check of the factor structure of the independent variables in Table 2 was made to see if the variables could be more parsimoniously described. Jöreskog and Sörbom's exploratory factor analysis was used (Jöreskog and Sörbom, 1978). Two of the variables, KDAYPRES and KUNEXABS, which measured days present and excused absences were collinear and KUNEXABS was dropped from the analysis. Examination of hypothesized factor analysis models showed successively increasing the factors improved the fit and that independent variables did not generally load on the same factor.

Description of the Gain Score and Longitudinal Models

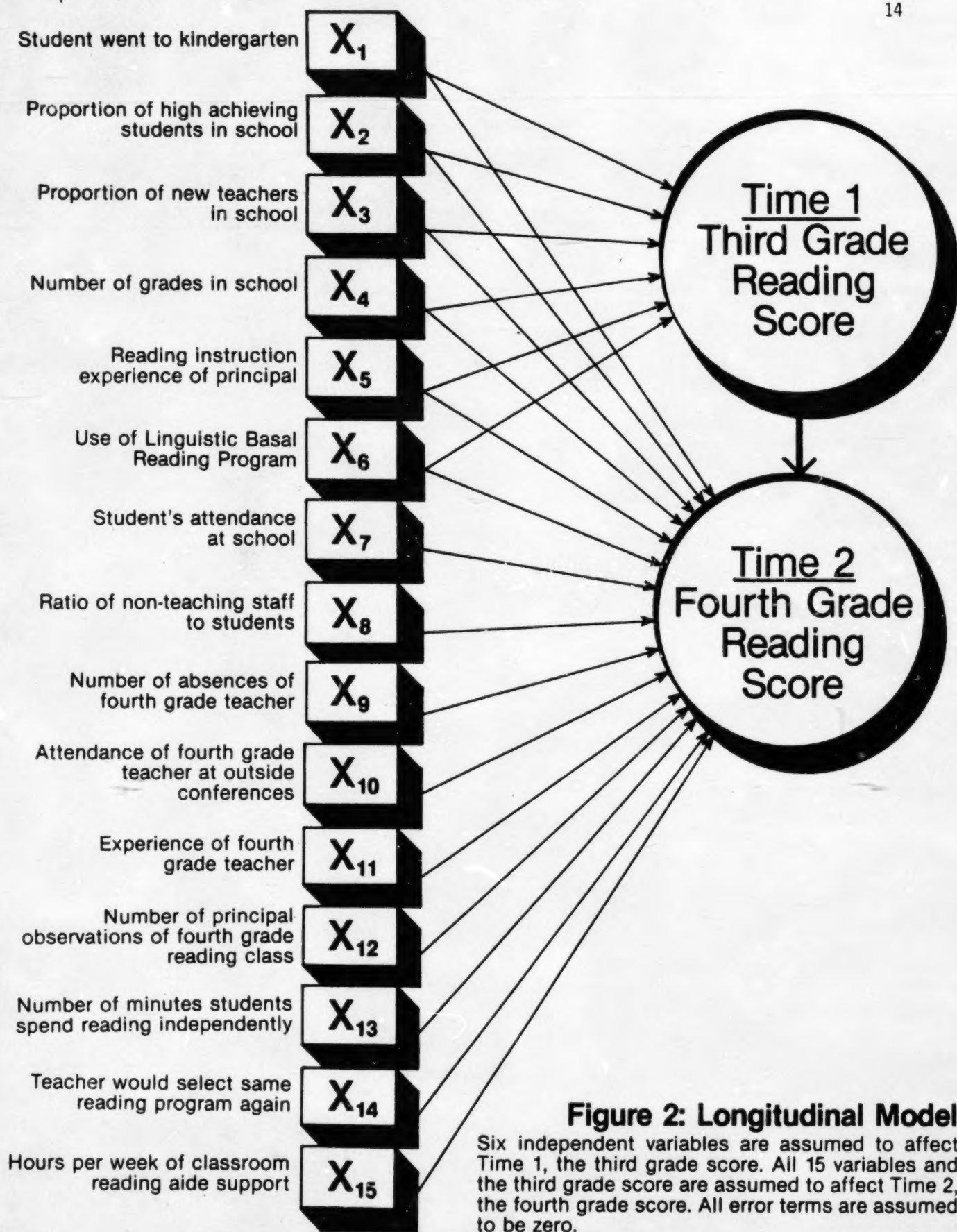
Figure 1 shows the gain score model. All 15 independent variables are assumed to influence the gain score. All variables are assumed to be measured without error. The gain score model was first analyzed using the step-wise multiple regression program of the Statistical Package for the Social Sciences (SPSS) and then using the just-identified multiple regression option of LISREL.



It is reasonable to ask if the use of a two-wave, one-variable model provides new information about the data base. Figure 2 shows one alternative model for analyzing the Philadelphia data. This model uses two dependent variables, the Time 1 and Time 2 variables, together instead of analyzing their difference. Two multiple regression runs were made using the 15 variables, first the third grade achievement variable was used as the dependent variable then the fourth grade variable was used.

A review of "WHAT WORKS IN READING?" and conversations with Philadelphia school district research and evaluation staff showed that 6 of the 15 variables can be hypothesized to influence both the third and fourth grade scores while the other 9 can be hypothesized to influence only the fourth grade score. The 6 variables influencing scores at both times were whether or not the student went to kindergarten (KINDERGN), the proportion of students in the school scoring well on the achievement test (XPCTGT84), the percent of teachers who had been in the school less than two years (XPCRTLT2), the span of grades in the school (SGRDORG7), whether or not the principal had previous reading experience (PBRDGEXP), and the use of a Linguistic Basal reading program (T55LNBSL).

Figure 2 contains two structural equations. The first uses the six independent variables affecting the third grade score. The second uses all 15 independent variables plus the third grade score. In this model, error terms are assumed to be zero. Equations comprising the maximum-likelihood analyses of covariances have been described elsewhere at length (see, for example, Long, 1976) and while detailed explication of them again here does not seem necessary, a brief overview may be of use to the reader.



Briefly, the LISREL model consists of two parts: the measurement model and the structural model. The measurement model specifies how latent or hypothetical constructs are measured in terms of observed variables. There are two measurement models, one for the observed dependent variables. Let $x' = (x_1, x_2, \dots, x_q)$ be a vector of observed independent variables and let $y' = (y_1, y_2, \dots, y_p)$ be a vector of observed dependent variables. Then,

$$\underline{x} = \underline{\Lambda}_x \underline{\xi} + \underline{\delta} \quad (1)$$

$$\underline{y} = \underline{\Lambda}_y \underline{\eta} + \underline{\varepsilon} \quad (2)$$

where $\underline{\xi}$ and $\underline{\eta}$ are random vectors of latent independent and dependent variables, $\underline{\xi}' = (\xi_1, \xi_2, \dots, \xi_n)$ and $\underline{\eta}' = (\eta_1, \eta_2, \dots, \eta_m)$ respectively. The vectors $\underline{\delta}$ and $\underline{\varepsilon}$ are errors of measurement in $\underline{y}$ and $\underline{x}$ respectively when $\underline{y}$ and $\underline{x}$ are measured as deviations from their means. The matrices $\underline{\Lambda}_x (q \times n)$ and $\underline{\Lambda}_y (p \times m)$ are regression matrices of ($\underline{x}$ on $\underline{\xi}$ and $\underline{y}$ on $\underline{\eta}$) respectively.

The structural model linking the two measurement models is given in (3).

$$\underline{\beta} \underline{\eta} = \underline{\Gamma} \underline{\xi} + \underline{\zeta} \quad (3)$$

Where $\underline{\beta}$ and $\underline{\Gamma}$ are coefficient matrices and $\underline{\zeta} = (\zeta_1, \zeta_2, \dots, \zeta_m)$ is a random vector of residuals reflecting disturbance terms or errors in equations (Jöreskog and Sörbom, 1978).

The two equations comprising the model in Figure 2 are given in (4) and (5).

$$\eta_1 = \underline{\Gamma}_1 \underline{\xi} + \zeta_1 \quad (4)$$

$$-\beta_{21} \eta_1 + \eta_2 = \underline{\Gamma}_2 \underline{\xi} + \zeta_2 \quad (5)$$

η_1 , the third grade score is seen to be comprised of two parts, $\underline{\Gamma}_1 \underline{\xi}$, which represents the effects of the six independent factors and ζ_1 which is the residual variance unexplained by the six variables affecting the third grade score. Equation (5) can also be rewritten as (6).

$$\eta_2 = \beta_{21} \eta_1 + \underline{\Gamma}_2 \underline{\xi} + \zeta_2 \quad (6)$$

η_2 , the fourth grade score is seen to be comprised of three parts: $\beta_{21}\eta_1$, which represents the effect of η_1 on η_2 ; $\Gamma_2\xi$, which is the effect of the 15 independent variables; and ζ_2 , which is the unexplained residual variance of η_2 .

This model is recursive in that the two dependent variables are not assumed to loop back and affect variables affecting them, e.g., the fourth grade score is assumed to have no effect on the third grade score. Identification problems in recursive models have been frequently commented on (Heise, 1969, 1970). In order to make recursive models identifiable, restrictive assumptions are generally made about error terms. For example, in order to identify the model in Figure 2, the usual procedure is to assume that all error terms and covariances of error terms are zero and uncorrelated with each other. This assumption includes assuming the covariance of the error terms, ζ_1 and ζ_2 in equations (4) and (5), is equal to zero.

The assumption of zero error terms and zero error covariances is equivalent to assuming that all variables are measured without error and the two measurement models are collapsed into the structural model. On the independent variable side of the model, this situation is called "fixed X"; Λ_x is assumed to be an identity matrix, I , the independent variable error matrix $\theta_\delta = 0$, and $x \equiv \xi$. Similar assumptions are made on the dependent variable side, i.e., $\Lambda_y = I$, $\theta_e = 0$ and $y \equiv \eta$. Given these assumptions, the structural equation model becomes the following.

$$\beta y = \Gamma x + \zeta \quad (9)$$

Table 4 compares the results of the multiple regression analyses with the structural equation analyses. For comparative purposes, both a correlation matrix and a covariance matrix were analyzed using LISREL. The computation of the covariance matrix requires both the correlations and standard

TABLE 4
Compares Multiple Regression and Structural Equations Analyses

	Statistical Package for Social Sciences (SPSS)			Structural Equation Analysis (LISREL) of Correlation Matrix			Structural Equation Analysis (LISREL) of Covariance Matrix	
	Gain Score Column 1	Time 1 Reading Score Column 2	Time 2 Reading Score Column 3	Gain Score Column 4	Time 1 Reading Score Column 5	Time 2 Reading Score Column 6	Time 1 Reading Score Column 7	Time 2 Reading Score Column 8
KDAY PRES	.072	----	.070	.066	----	.075	----	.523
KINDERGN	.024	.046	.030	.019	.068	.057	11.304	6.208
SUPSTFPK	-.177	----	.132	-.107	----	.081	----	.536
XPCTGT84	-.196	.452	.206	-.114	.383	.278	199.304	85.907
XPCRTLT2	.123	-.228	-.066	.029	-.053	.012	-26.610	15.843
SGRDPRG7	-.052	.116	.033	-.020	.045	.020	6.587	.816
TPPNDBS	.068	----	.029	.062	----	.033	----	.634
TATTCNF	-.059	----	.012	-.058	----	.016	----	2.955
TISTGR04	*	----	-.059	-.017	----	-.047	----	-9.727
PBROGEXP	.086	.018	-.074	.106	-.053	-.113	-7.616	-14.722
TPROPSRP	-.042	----	.018	-.040	----	.022	----	.136
TMINORDG	.042	----	.123	.040	----	.126	----	.152
TSELRDGM	-.006	----	.042	-.007	----	.041	----	5.918
T55LNBSL	.055	.048	.111	.087	.000	.086	.065	21.559
TNAIDEMK	.038	----	-.107	.042	----	.107	----	-3.357
KCATGR03	----	----	.599	----	----	.288	----	.656
R ² or 1-ψ	.049	.20	.57	.043	.16	.48	.18	.47
χ^2_{df}					$\chi^2_{df} = \frac{224.82}{9} = 24.9$		$\chi^2_{df} = \frac{266.04}{9} = 29.6$	

*F-level or tolerance level insufficient for further computation.

TABLE 4

deviations of the variables. As shown in Table 4, the use of the correlation and covariance analyses yields different results for the estimation of the effect that the third grade score, KCATGRD3, has on the fourth grade variable, KCATGRD4. Consider the following two cases: In Case 1, students spread out more between Time 1 and Time 2. Perhaps the better students get better faster. The variance at Time 1 might be 1.00 and the variance at Time 2 = 2.00. In case 2, the spread between the times is equal at all levels, or students could be closer together, so that Time 1 variance $\geq$ Time 2 variance.

Using a correlation matrix explains the relative variance; i.e., where students are in relation to one another, but does not account for absolute change. The use of a covariance matrix creates one underlying scale on which both variables can be placed and student movement on the underlying scale can be measured. Thus, a covariance analysis presumably yields a better estimate of the effect of the third grade score on the fourth.

The first column in Table 4 shows the multiple regression results of using all 15 variables against the gain score, CATGAIN. The second column shows the results of running the 6 variables against the third grade score, again using a multiple regression approach, and the third column shows the multiple regression results when all 15 variables plus the third grade variable are run against the fourth grade variable. The multiple SPSS regression approach accounted for .049 of the variance in the gain score, .195 of the third grade score, and .569 of the fourth grade score. The beta coefficient estimating the effect of the third grade score on the fourth grade score was .599. These first three columns show systematic differences in the signs of the beta coefficients. All of the coefficients

with negative signs in the gain score analysis have positive signs when the components of the gain score are considered separately. The four variables whose coefficients have negative signs in the second and third columns have positive signs when run against the gain score only. The changes of sign can create difficulty in the interpretation of results. For example, a review of interpretation of the Philadelphia data indicates several variables whose directions of influence on reading achievement are unclear.

Columns 4, 5, and 6 of Table 4 show the structural equation results obtained analyzing a correlation input matrix. Column 4 shows the results when all 15 independent variables are run against the gain score. The signs of the structural coefficients are similar to the signs obtained in the SPSS step-wise multiple regression, however, the magnitudes of the coefficients are generally lower. Columns 5 and 6 present the results of analyzing the two structural equations. Again, the structural coefficients are similar in sign and generally smaller in magnitude than the multiple regression coefficients. The structural coefficient estimate of the effect of the third grade score on the fourth, β_{21} , is .288, approximately .30 less than the SPSS multiple regression least squares estimate.

As the bottom of columns 4, 5, and 6 respectively indicate, the LISREL correlation analysis accounted for .043 of the variance in the gain score, CATGAIN; .16 of the variance in the Time 1 score and .48 of the Time 2 variance.

Columns 6, 7, and 8 present the structural equation analysis of the covariance matrix. In the covariance analysis, the structural coefficients are unstandardized and reported in the same units used to describe the variables. The covariance analysis shows that .18 of the variance in the

Time 1 score is accounted for by the six independent variables run against it, and .48 of the variance in the Time 2 score is accounted for by all 15 independent variables and the Time 1 score.

A major difference between the two LISREL approaches versus the SPSS multiple regression is in the analysis of Time 2. The LISREL approaches account for approximately .48 of the variance in Time 2 while the SPSS multiple regression approach accounts for .57.

A major difference between the LISREL correlation results versus the LISREL covariance results is the calculation of, β_{21} , the effect of Time 1, η_1 on Time 2, η_2 . The SPSS multiple regression analysis estimates β_{21} at .599, the correlation analysis estimates it at .288, and the covariance analysis estimates it at .656. Since the covariance matrix uses more information about the variables than the correlation matrix, the covariance approach is preferable on theoretical grounds. The covariance matrix is also preferable because of its use of unstandardized variances and coefficients. Examples showing the merit of unstandardized coefficients include Schoenberg, 1972; Sprecht and Warren, 1976; Kim and Mueller, 1976; and Woefel and Woefel, 1977.

The bottom row of Table 4 shows the maximum likelihood goodness of fit statistics which indicate how well the hypothesized correlation and covariance analyses of the model in Figure 2 fits the data. A χ^2 goodness-of-fit test is used to determine the extent to which estimated parameters can reproduce the original input data matrix. Values of χ^2 which are large relative to the degrees of freedom indicate a poor fit. χ^2 values approaching the degrees of freedom indicate a better fit. (Long, 1976: 168-173; Jöreskog and Sörbom, 1977: 290-291.) The bottom row of Table 4

shows the ratio of the χ^2 to the degrees of freedom for both the correlation and the covariance specifications. While the correlation fit ratio is better than the covariance fit ratio, both ratios are ten times larger than acceptable. These results indicate that the model in Figure 2 needs to be revised. No goodness of fit test is available for the gain score results since multiple regression is a just-identified model containing zero degrees of freedom and zero chi-square.

COVARIANCE MODELS INCLUDING ERROR ESTIMATIONS

The Need to Include Estimations of Error

As described above, typical identifications of two-wave one-variable models assume no errors exist in the measurement of variables, no error covariances exist and the disturbances of the equations are uncorrelated. Bohrnstedt and Carter's (1971) and Alvin and Jackson's (1980) admonitions that path analysis is based on very restrictive assumptions and that such assumptions reflect a blatant unconcern with measurement error problems are also pertinent here.

The error assumptions typically made in order to identify two-wave one-variable models are unrealistic. For example, in the Philadelphia data it is reasonable to assume there is some measurement error in the measurement of third and fourth grade reading achievement. Since a similar measuring instrument was used at both times, it may be reasonable to hypothesize correlated measurement error (see Wiley and Wiley, 1974). Moreover, the Philadelphia analysis implicitly assumes that CATGAIN is measured without error. The fifteen independent variables probably contain some measurement

error. The effect of omitted variables on the studied set is probably greater than zero. Moreover, given the low percentage of variance accounted for in the dependent variables and the over 250 variables contained in the data base of "WHAT WORKS IN READING?" it is reasonable to assume the errors of the two structural equations may be correlated.

Part Two of this paper discusses alternative models including measurement error. After discussing methods of further reducing the number of independent variables studied, the results of an exploratory factor analysis of 11 independent variables are described. Alternative models of the Philadelphia data are systematically analyzed using a "sensitivity" analysis to consider the effect of various error assumptions.

END OF PART ONE

FOOTNOTES

1. The sample of 1,828 students was divided into an "experimental" sample of 1,363 and a "cross-validation" sample of 465. "WHAT WORKS IN READING?" has two tables in it. The first table shows the unstandardized beta weights of 35 variables for both samples, shows t-ratio for both coefficients, and shows the results of a test of equality of coefficients.

2. The maximum likelihood estimates are not unbiased (Hogg and Craig, 1970). Unbiased estimates correspond to the arithmetic mean and maximum likelihood estimates to the mode (Tintner, 1968).

3. A complete correlation matrix for all independent variables and the three dependent variables for the 1,828 students could not be obtained. A complete correlation matrix for 16 independent variables and the three dependent variables was obtainable for the experimental sample of 1,363 students. All variable labels and correlations were verified by research and evaluation staff of the Philadelphia school district.

4. The third and fourth grade scores and their correlation with CATGAIN are not considered here because of the difficulties of interpreting correlations of ratios with common terms (Fuguitt and Lieberman, 1974).

REFERENCES

- ALWIN, D. F. and D. J. JACKSON (1980) "Measurement Models for Response Errors in Surveys: Issues and Applications," pp. 68-119 in K. F. Schuessler (ed.) *Soc. Methodology* 1980. San Francisco: Jossey-Bass.
- ALWIN, D. F. and M. J. SULLIVAN (1975) "Issues of Design Design and Analysis in Evaluation Research." *Soc. Methods and Research* 4 (August): 77-100.
- BLALOCK, H. M. (1964) *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.
- BOHRNSTEDT, G. W. (1969) "Observations on the measurement of change," pp. 113-133 in E. F. Borgatta and G. W. Bohrnstedt (eds.) *Soc. Methodology* 1969. San Francisco: Jossey-Bass.
- BOHRNSTEDT, G. W. and T. M. CARTER (1971) "Robustness in Regression Analysis," pp. 118-146 in H. L. Costner (ed.) *Sociological Methodology* 1971. San Francisco: Jossey-Bass.
- BOUDON, R. (1965) "A Method of Linear Causal Analysis: Dependence Analysis." *Amer. Soc. Rev.* 30: 365-374.
- BURT, R. S. (1973) "Confirmatory Factor-Analytic Structures and the Theory Construction Process." *Soc. Methods and Research* 2: 131-190.
- COSTNER, H. L. (1969) "Theory, Deduction, and Rules of Correspondence." *Amer. J. of Sociology* 75: 245-263.
- CRONBACH, L. J. and L. FURBY (1970) "How We Should Measure 'Change'--or Should We?" *Psychological Bulletin* 74: 68-80.
- DUNCAN, C. D. (1972) "Unmeasured Variables in Linear Models" pp. 36-82 in H. L. Costner (ed.) *Sociological Methodology* 1972, pp. 36-82. San Francisco: Jossey-Bass.

- DUNCAN, O. D. (1969) "Some linear models for two-wave, two-variable panel analysis." *Psychological Bulletin* 77: 177-182.
- DUNCAN, O. D. (1966) "Path Analysis: sociological examples." *Amer. J. of Sociology* 72: 1-16.
- FINK, E. L. and T. I. MABEE (1978) "Linear Equations and non-linear estimations." *Soc. Methods and Research* 7 (August): 107-120.
- FISHER, R. A. (1921) "On the mathematical foundations of theoretical statistics." *Phil. Transactions A*. 222: 309.
- FUGUITT, G. V. and S. LIEBERSON (1974) "Correlation of Ratios or Difference Scores Having Common Terms," pp. 128-144 in H. L. Costner (ed.) *Sociological Methodology 1973-1974*. San Francisco: Jossey-Bass.
- GILLESPIE, M. G. and J. FOX (1980) "Specification Error and Negatively Correlated Disturbances in "Parallel Simultaneous-Equation Models." *Soc. Methods and Research* 8: 273-308.
- HANNAN, M. T. and A. A. YOUNG (1977) "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series," pp. 52-83 in D. R. Heise (ed.) *Sociological Methodology 1977*. San Francisco: Jossey-Bass.
- HARGENS, L., B. F. RESKIN, and P. D. ALLISON (1976) "Problems in Estimating Measurement Error from Panel Data," *Soc. Methods and Research* 4 (May): 439-458.
- HAUSER, R. M. and A. GOLDBERGER (1971) "The Treatment of Unobservable Variables in Path Analysis," pp. 81-117 in H. Costner (ed.) *Sociological Methodology 1971*. San Francisco: Jossey-Bass.
- HEISE, D. R. (1970) "Causal Inference from Panel Data," pp. 3-27 in E. B. Borgatta (ed.) *Sociological Methodology 1970*. San Francisco: Jossey-Bass.

- ___ (1969) "Problems in Path Analysis and Causal Inference," pp. 38-73 in
E. F. Borgatta (ed.) Sociological Methodology 1969. San Francisco:
Jossey-Bass.
- HOGG, R. V. and A. T. CRAIG (1970) Introduction to Mathematical Statistics.
New York: MacMillan Co.
- HORN, J. L. and J. J. McARDLE (1980) "Perspectives on Mathematical/
Statistical Model Building (MASMOB) in Research on Aging," pp. 503-541
in L. W. Poan (ed.) Aging in the 1980s: Psychological Issues.
Washington, D. C.: American Psychological Association.
- JORESKOG, K. G. and D. SORBOM (1978) EFAP-II Exploratory Factor Analysis
Program. Uppsala, Sweden: University of Uppsala.
- ___ K. G. and D. SORBOM (1978) LISREL-IV Analysis of Linear Structural
Relationships by the Method of Maximum Likelihood. Chicago:
International Educational Services.
- ___ and D. SORBOM (1977) "Statistical Models and Methods for Analysis of
Longitudinal Data," pp 285-325 in D. V. Aigner and A. S. Goldberger
(eds.) Latent Variables in Socioeconomic Models. Amsterdam: No. Holland
Publishing Co.
- KEAN, M. H., A. S. SUMMERS, M. J. RAIVETZ, and I. J. FARBER (1979) WHAT
WORKS IN READING?: The Results of a Joint School District/Federal
Reserve Bank Empirical Study in Philadelphia (May) Office of Research
and Evaluation, The School District of Philadelphia.
- ___ (1979) WHAT WORKS IN READING?: Technical Supplement of a Joint School
District/Federal Reserve Bank Empirical Study in Philadelphia.
(November) Office of Research and Evaluation, The School District of
Philadelphia.

- KESSLER, R. C. (1977) "The use of change scores as criteria in longitudinal survey research." *Quality and Quantity* 11: 43-66.
- KIM, J. and C. W. MUELLER (1976) "Standardized and Unstandardized Coefficients in Causal Analysis: An Expository Note." *Soc. Methods and Research* 4 (May): 423-437.
- LAND, K. C. (1969) "Principles of Path Analysis," pp. 3-37 in E. F. Borgatta (ed.) *Sociological Methodology*. San Francisco: Jossey-Bass.
- ____ and M. FELSON (1978) "Sensitivity Analysis of Arbitrarily Identified Simultaneous-Equation Models." *Soc. Methods and Research* 6 (February): 283-307.
- LONG, J. S. (1976) "Estimation and Hypothesis Testing in Linear Models Containing Measurement Error: A Review of Joreskog's Model for the Analysis of Covariance Structures." *Soc. Methods and Research* 5 (November): 157-206.
- PENDLETON, B. F., R. D. WARREN, and H. C. CHANG (1979) "Correlated Denominators in Multiple Regression and Change Analysis." *Soc. Methods and Research* 7: 451-474.
- SARIS, W. E., DE PIJPER, W. M. and P. ZEGWAART (1979) "Detection of Specification Errors in Linear Structural Equation Models," in Karl F. Schuessler (ed.) *Sociological Methodology* 1979, pp. 151-171. San Francisco: Jossey-Bass
- SCHMIDT, W. H. and V. SCHIEFLEY (1980) "An Empirical Examination of the Properties of Maximum Likelihood procedures for Covariance Structure Models." Paper presented at Annual Meeting of American Educational Research Association. Boston.

- SCHOENBERG, R. (1972) "Strategies for Meaningful Comparison," pp. 1-35
in H. Costner (ed.) *Sociological Methodology 1972*. San Francisco:
Jossey-Bass.
- SIMON, H. (1957) *Models of Man*. New York: Wiley.
- SPRECHT, D. A. and R. A. WARREN (1976) "Comparing Causal Models," pp. 46-82
in D. R. Heise (ed.) *Sociological Methodology 1976*. San Francisco:
Jossey-Bass.
- THORNDIKE, R. L. (1966) "Intellectual Status and Intellectual Growth."
Journal of Educational Psychology 57: 121-127.
- ___ (1963) *The Concepts of Over and Under Achievement*. New York: Bureau
of Publications, Columbia University.
- ___ and E. HAGEN (1955) *Measurement and Evaluation in Psychology and
Education*. New York: John Wiley.
- TINTER, G. (1968) *Methodology of Mathematical Economics and Econometrics*.
Chicago: University of Chicago Press.
- WHEATON, B., B. MULTHEN, D. ALWIN, and G. F. SUMMERS (1977) "Assessing
Reliability and Stability in Panel Models," pp. 84-136 in D. R. Heise
(ed.) *Sociology Methodology 1977*. San Francisco: Jossey-Bass.
- WILEY, D. E. and J. A. WILEY (1970) "The Estimation of Measurement Error in
Panel Data." *Amer. Soc. Rev.* 35: 112-116.
- WILEY, J. A. and M. G. WILEY (1974) "A Note on Correlated Errors in
Repeated Measurement." *Soc. Methods and Research* 3: 172-188.
- WOELFEL, J. C., WOELFEL, J. D. and M. L. WOELFEL (1977) "Standardized
versus Unstandardized Data Matrices: Which Type is More Appropriate for
Factor Analysis." Paper presented at Annual Meeting of American
Sociological Association (September 1977).